stochastic
processes
and their

www.elsevier.com/locate/spa

A statistically and computationally efficient method for frequency estimation

Kai-Sheng Song[a,*], Ta-Hsin Li[b,1]

[a] Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, USA
[b] Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA

Received 24 August 1998; accepted 6 September 1999

Abstract

Traditional methods of estimating frequencies of sinusoids from noisy data include periodogram maximization and nonlinear least squares, which lead to efficient estimates with the rate $\mathcal{O}(n^{-3/2})$. To actually compute these estimates, some iterative search procedures have to be employed because of the high nonlinearity in the frequency parameters. The presence of many local extrema requires the search be started with a very good initial guess – the required precision is typically $o(n^{-1})$, which is not readily available even from the fast Fourier transform of the data. To overcome these problems, we consider an alternative approach, the contraction-mapping (CM) method. Contributions of this paper include: (a) the establishment, for the first time, of the crucial connection between the accuracy of the initial guess required for convergence in the fixed-point iteration and the precision of the CM estimator as the fixed point of the iteration; (b) the quantification of the asymptotic relationship between the initial guess and the final CM estimator, together with limiting distributions and almost sure convergence of the fixed point; and (c) the construction of a single algorithm adaptable to possibly poor initial values without requiring separate procedures to provide initial guesses. It is shown that the CM algorithm, endowed with an adaptive regularization parameter, can accommodate possibly poor initial values of precision $\mathcal{O}(1)$ and converge to a final estimator whose precision is arbitrarily close to the optimal $\mathcal{O}(n^{-3/2})$.

Keywords: Filter; Fixed points; Frequency estimation; Martingale difference; Nonlinear regression; Signal processing; Spectral analysis; Time series

1. Introduction

Let $\{y_1, \ldots, y_n\}$ be a time series of length n obtained from a zero-mean stationary random process of the form

$$y_t = \beta \cos(\omega_0 t + \phi) + \varepsilon_t, \tag{1.1}$$

* Corresponding author.

E-mail addresses: kssong@stat.fsu.edu (K.-S. Song), thl@us.ibm.com (T.-H. Li)

[1] Now with the Department of Mathematical Sciences, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598-0218, USA

PII: S0304-4149(00)00085-X

where $\beta > 0$ and $\omega_0 \in (0, \pi)$ are unknown constants, ϕ is a uniformly distributed random variable in $(-\pi, \pi]$, and $\{\varepsilon_t\}$ is a zero-mean stationary ergodic process which is independent of ϕ. In this paper we consider the problem of estimating the unknown frequency ω_0 on the basis of the observed data record $\{y_1, \dots, y_n\}$, i.e., we focus on estimation of single hidden frequency. For generalization of our results to estimation of more than one hidden frequency (i.e., multiple frequencies in a model of superposition of a finite number of sinusoids in stationary noise), see the concluding remarks in Section 6.

The problem of frequency estimation has been an important subject in statistical signal processing and time series analysis (e.g., Kay, 1988; Priestley, 1981). Traditional methods of frequency estimation include periodogram maximization (PM) (Walker, 1971) and nonlinear least squares (NLS) (which is maximum likelihood if $\{\varepsilon_t\}$ is Gaussian) (Hannan, 1971; Rao and Zhao, 1993). Although these methods lead to efficient frequency estimates whose asymptotic standard error is of order $\mathcal{O}(n^{-3/2})$, it is not a trivial task to actually compute the estimates. Because the problems are highly nonlinear, iterative search procedures are usually required. However, the presence of many spurious local extrema in the vicinity of the solutions makes the iteration extremely sensitive to the starting values – the required precision of the starting values is typically $o(n^{-1})$, which is not readily available even from the fast Fourier transform (FFT) of the data, whose precision for frequency estimation is merely $\mathcal{O}(n^{-1})$ (Rice and Rosenblatt, 1988). To overcome the difficulty, heuristic methods are typically employed that improve upon the FFT estimates and thus produce the desired initial values (e.g., Stoica et al., 1989). To fully understand the problem of frequency estimation from both theoretical and practical viewpoints, the issue of initial values must be taken into consideration. In the existing literature, although very huge indeed, the initial value problem has not been adequately addressed. This leaves a great gap between theory and practice. This paper addresses this issue, among others, by establishing a rigorous, unified theory and thus fills up the gap on a sound basis.

An alternative approach for frequency estimation is called iterative filtering (IF) (e.g., Dragoševiȁ and Stankoviȁ, 1989; Kay, 1984; Nehorai, 1985; Stoica and Nehorai, 1988; Tichavský and Händel, 1995). This approach is favored in many engineering applications because of its effectiveness and simplicity. In particular, the filtering in most IF methods can be computed recursively, so, unlike FFT, the resulting frequency estimators can be easily updated upon the arrival of new observations without having to reprocess the entire time series (e.g., Li and Kedem, 1998). The *contraction mapping* (CM) method proposed by Li and Kedem (1993a) is based on the IF approach. The CM method alternates the following steps: (a) pass the observed time series through a parametric filter whose parameter is determined by the previous frequency estimate; (b) compute the lag-one autocorrelation coefficient of the output time series to obtain an improved frequency estimate. As such, the CM method can be unified under the general concept of *parametric filtering* (PF) which advocates the combination of linear parametric filters with simple output summary statistics for time series analysis (Kedem, 1994; Li, 1996,1997).

The parametric filter employed by Li and Kedem (1993a) is a second-order autoregressive (AR) filter endowed with an extra regularization parameter $\eta \in (0, 1)$ that

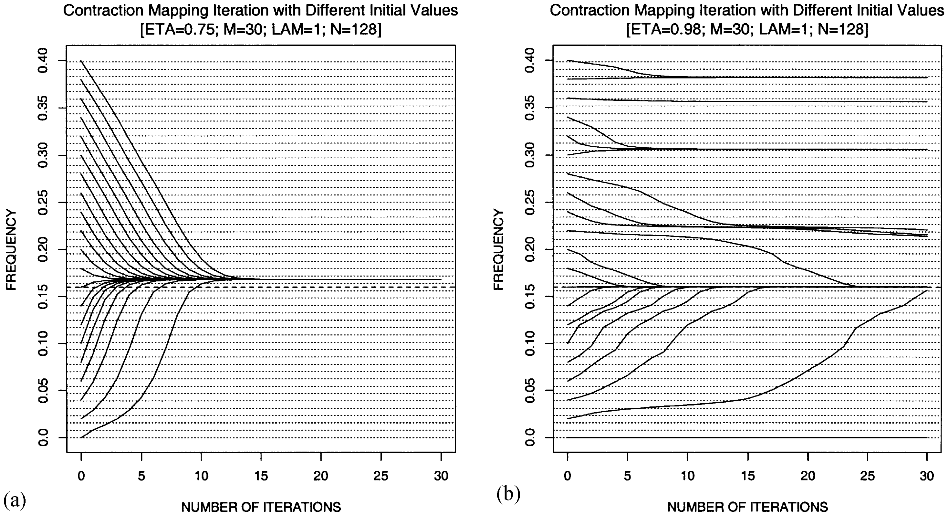

Fig. 1. Trajectory of the CM iterations with different initial values. The solid lines represent the CM frequency estimates $\hat{f}_n^{(m)} := \hat{\omega}_n^{(m)}/(2\pi)$ plotted as functions of the iteration number m, the dashed line indicates the true frequency $f_0 := \omega_0/(2\pi)$, and the dotted lines stand for the Fourier frequencies separated by n^{-1}. (a) With a small $\eta = 0.75$, the algorithm is less sensitive to initial values for convergence, but the final frequency estimates are less accurate. (b) With a large $\eta = 0.98$, the algorithm is more sensitive to initial values for convergence to the desired solution, but if converges the frequency estimates are more accurate. In this example, $n = 128$, $f_0 = 20.5\, n^{-1} = 0.1601562$, and $\phi = 0.2\pi$. The noise $\{\varepsilon_t\}$ is white Gaussian with the signal-to-noise ratio equal to -3 dB.

stabilizes the filter by restricting its poles (the conjugate pair of roots of the second-degree AR polynomial) outside the unit circle of the complex plane. An important benefit of the regularization parameter is the relaxation of the requirements on initial values. Indeed, it can be shown that with a suitable choice of η the CM method allows poor initial values of precision $\mathcal{O}(1)$ and still manages to converge to the desired solution.

Numerical experiments show that the accuracy of the CM estimator depends not only on n but also on η – the closer is η to unity the more accurate is the resulting frequency estimate. However, as η becomes near unity, more accurate initial values are needed for the CM iterations to converge. Fig. 1 presents a simulated example that demonstrates this phenomenon. Therefore, to successfully implement the CM method in practice, one should start with a small η to accommodate possibly poor initial values and then gradually increase η toward unity as improved frequency estimates from previous iterations become accurate enough to initiate the iteration with the increased η. In so doing, the CM method can start with poor initial values of precision $\mathcal{O}(1)$ and produce, after the iterations, highly accurate frequency estimates whose precision can be arbitrarily close to $\mathcal{O}(n^{-3/2})$. This idea is illustrated by the example in Fig. 2.

Simulation also shows that when η becomes too close to unity relative to the sample size n, the asymptotic results of Li and Kedem (1993a), Li et al. (1994) are no longer valid because they are derived under the scenario (assumptions) that η is constant while n approaches infinity. A more appropriate analysis requires η to be adaptive to the sample size n and approach unity at a certain rate as n tends to infinity. Along these lines,

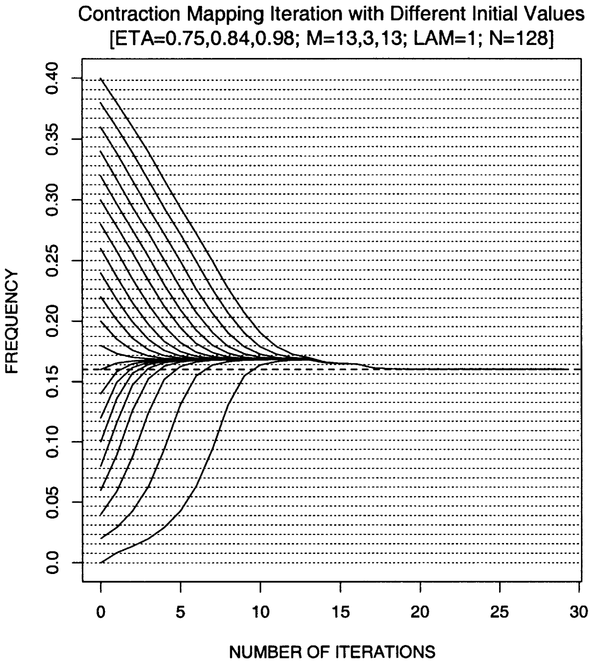

Fig. 2. Similar to Fig. 1, except that η is increased from 0.75 to 0.84 after 13 iterations and then to 0.98 after 3 more iterations. Unlike Fig. 1(b), all initial values lead to the desired solution, and unlike Fig. 1(a), the final frequency estimates are very accurate.

we provide, in this paper, a rigorous theoretical analysis that vindicates the empirical findings discussed before and extends the asymptotic studies of Li and Kedem (1993a) and Li et al. (1994). One of the important contributions of this analysis is the analytical formulas that provide refined quantitative relationships between the requirement of initial values for the CM algorithm to converge and the asymptotic standard error of the resulting frequency estimates, all in terms of the closeness of η to unity which now is a function of n instead of a constant. These theoretical results provide a useful guide for the selection of η in practice as demonstrated by the simulation examples in Section 4.

Given that $\mathcal{O}(n^{-1/2})$ is the accuracy in most parametric estimation problems and seems quite satisfactory, one cannot help but wonder why frequency estimation requires an accuracy which is not only better than $\mathcal{O}(n^{-1/2})$ but $\mathcal{O}(n^{-1})$. This question can be answered by considering the problem of frequency detection, i.e., that of deciding whether or not a sinusoid of unknown frequency is present in the data. Typical methods of frequency detection rely upon an accurate assessment of the sinusoid's amplitude, by linear regression, for example, once the frequency becomes available. However, the amplitude cannot be estimated consistently unless the frequency is estimated with a precision of $o(n^{-1})$ or better (e.g., Rice and Rosenblatt, 1988). Otherwise one would face an error-in-variable problem when estimating the amplitude on the basis of the inaccurate estimated sinusoid (i.e., the explanatory variable measured with error in the terminology of linear regression) in which the resulting error (i.e., error in the

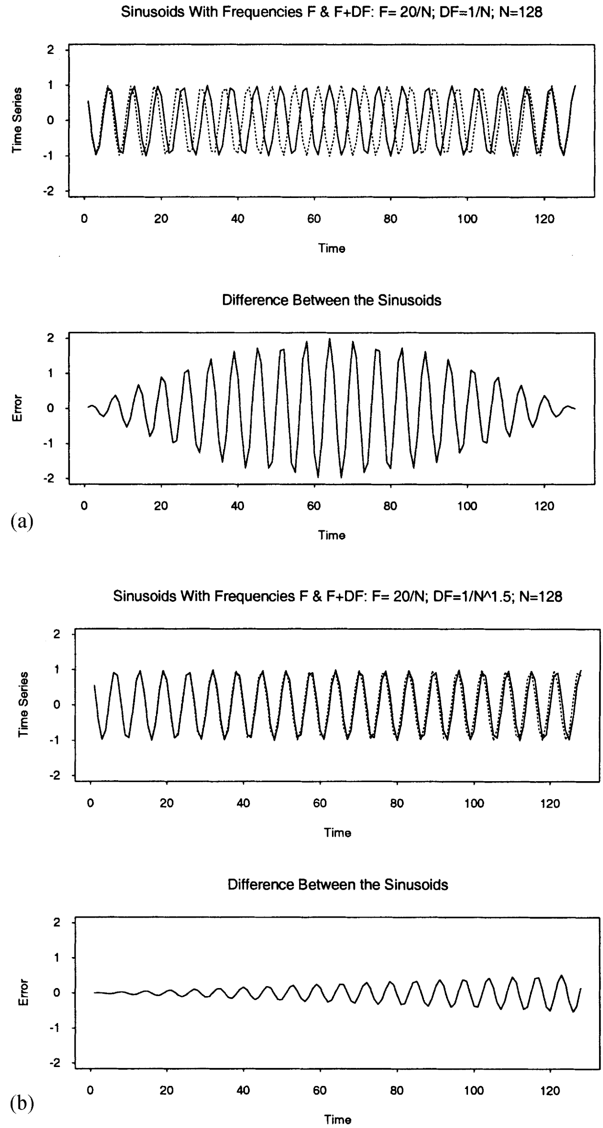

Fig. 3. Error-in-variable problem in frequency detection. (a) An estimated sinusoid (dotted line) whose frequency differs from the true value by n^{-1}; (b) An estimated sinusoid whose frequency differs from the true value by $n^{-3/2}$. Solid lines in the first row represent the sinusoids with the exact frequency. The resulting errors (i.e., errors in the explanatory variable) are shown in the second row. In both cases $n = 128$.

explanatory variable) can be as large as $\mathcal{O}(1)$, so the amplitude estimator is inconsistent. Fig. 3 illustrates this point.

The rest of the paper is organized as follows. Section 2 describes the CM method and summarizes the results of Li and Kedem (1993a) and Li et al. (1994). Section 3 presents the main results of the paper concerning (a) the requirement of initial values that ensure the convergence of the CM iterations, (b) the asymptotic distributions of the

CM estimators under different assumptions about η and (c) the construction of a single unified algorithm that can accommodate possibly poor initial values of precision $\mathcal{O}(1)$ and converge to a final estimator whose precision is arbitrarily close to the optimal rate $\mathcal{O}(n^{-3/2})$. Section 4 provides the results from some simulation experiments which intend to validate the asymptotic analysis. The theorems are proved in Section 5 based on some preliminary results whose proofs are not given in this paper since they are extremely lengthy and technical. For detailed proofs of these propositions and technical lemmas, see the technical report of Song and Li (1997). Section 6 ends with some concluding remarks.

2. The CM frequency estimator

For any $\eta \in (0,1)$ and $\alpha := \cos\omega \in \mathscr{A} := (-2\eta(1+\eta^2)^{-1}, 2\eta(1+\eta^2)^{-1})$, let $\{y_t(\alpha)\}$ be obtained recursively from the observed time series $\{y_1,\ldots,y_n\}$ according to

$$y_t(\alpha) + 2\theta(\alpha)\eta y_{t-1}(\alpha) + \eta^2 y_{t-2}(\alpha) = y_t \quad (t=1,\ldots,n), \tag{2.2}$$

with $y_{-1}(\alpha) = y_0(\alpha) := 0$. In other words, with $\{y_t\}$ as the input, the series $\{y_t(\alpha)\}$ is the output from the second-order autoregressive (AR) filter $\mathscr{L}(\mathscr{B};\alpha) = (1+2\theta(\alpha)\eta\mathscr{B} + \eta^2\mathscr{B}^2)^{-1}$, where $\mathscr{B}$ represents the backward-shift operator such that $\mathscr{B}y_t = y_{t-1}$. In the following, the parameter $\theta(\alpha)$ is always taken to be (Li and Kedem, 1993a)

$$\theta(\alpha) := -\frac{1+\eta^2}{2\eta}\alpha := -\cos\lambda. \tag{2.3}$$

Note that the variable $\lambda \in (0,\pi)$ always exists and is uniquely determined by η and α from (2.3) for any $\eta \in (0,1)$ and $\alpha \in \mathscr{A}$.

Let the lag-one autocorrelation coefficient of the filtered process $\{y_t(\alpha)\}$ be estimated by the least-squares estimator

$$\rho_n(\alpha) = \frac{\sum_{t=1}^n y_{t-1}(\alpha)\{y_t(\alpha) + \eta^2 y_{t-2}(\alpha)\}}{(1+\eta^2)\sum_{t=1}^n y_{t-1}^2(\alpha)}, \tag{2.4}$$

which minimizes the weighted sum of forward and backward prediction error sums of squares

$$e_n^2(\rho) := \sum_{t=1}^n \{y_t(\alpha) - \rho y_{t-1}(\alpha)\}^2 + \eta^2 \sum_{t=1}^n \{y_{t-2}(\alpha) - \rho y_{t-1}(\alpha)\}^2.$$

The contraction mapping (CM) method proposed by Li and Kedem (1993a) estimates the frequency ω_0 in (1.1) using the following iterative procedure:

$$\hat{\alpha}_n^{(m)} := \rho_n(\hat{\alpha}_n^{(m-1)}) \quad (m = 1,2,\ldots). \tag{2.5}$$

Suppose that with a suitable initial guess $\hat{\alpha}_n^{(0)}$ the CM iteration (2.5) converges to $\hat{\alpha}_n$, i.e., $\hat{\alpha}_n^{(m)} \to \hat{\alpha}_n$ as $m \to \infty$. Then, a frequency estimator is obtained from the transformation

$$\hat{\omega}_n := \arccos(\hat{\alpha}_n). \tag{2.6}$$

Clearly, one can regard $\hat{\alpha}_n$ as an estimator of the parameter $\alpha_0 := \cos\omega_0$, and regard $\hat{\omega}_n^{(m)} := \arccos(\hat{\alpha}_n^{(m)})$ as intermediate frequency estimates.

Some convergence properties of the CM algorithm (2.5) and asymptotic properties of the CM frequency estimator (2.6) were investigated by Li and Kedem (1993a) and Li et al. (1994) under the assumption that η is a constant and is independent of the sample size n. It was shown that for sufficiently large n the sequence $\{\hat{\alpha}_n^{(m)}\}$ converges to some $\hat{\alpha}_n$ almost surely as $m \to \infty$ if the iteration is initiated by any initial value $\hat{\alpha}_n^{(0)}$ such that $|\hat{\alpha}_n^{(0)} - \alpha_0| \leqslant c$ for some constant $c > 0$. It was also shown that the resulting CM frequency estimator $\hat{\omega}_n$ converges to ω_0 almost surely as $n \to \infty$ and is asymptotically normal with the standard error $((1-\eta^2)/(1+\eta^2))^{3/2}\gamma^{-1}n^{-1/2}$, where $\gamma := \frac{1}{2}\beta^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio of $\{y_t\}$ and σ_ε^2 is the variance of $\{\varepsilon_t\}$.

This expression of standard error suggests that improved accuracy can be achieved by making η close to unity. However, simulation shows that when η is too close to unity relative to the sample size, the foregoing standard error expression is no longer accurate (Li et al., 1994). Furthermore, when η is too close to unity relative to the sample size, the initial precision of $\mathcal{O}(1)$ becomes insufficient to ensure convergence. One of the main objectives of this paper is to study the convergence and asymptotic properties of the CM method with variable regularization parameter η adaptive to the sample size.

3. Main results

For any $\eta \in (0,1)$, one can always write

$$\eta = 1 - \delta, \quad \delta \in (0,1). \tag{3.1}$$

In the following, unless stated otherwise, we always assume that η, and hence δ, are functions of the sample size n such that $\eta \to 1$, or equivalently, $\delta = 1 - \eta \to 0$, as $n \to \infty$. For technical reasons, we also assume that $\{\varepsilon_t\}$ is a martingale difference sequence with respect to a filtration $\{\mathscr{F}_t\}$ such that $E\{\varepsilon_t^2|\mathscr{F}_{t-1}\} = \sigma_\varepsilon^2$ almost surely and $E\{\varepsilon_t^4\} < \infty$.

Theorem 1. *Let $\mathscr{A}_n := \{\alpha\colon |\alpha - \alpha_0| \leqslant \kappa(1-\eta)^\varepsilon\} \subset \mathscr{A}$ be a neighborhood of α_0, where $\kappa > 0$ and $\varepsilon \in (1, \frac{3}{2})$ are constants, and $\eta \in (0,1)$ is a function of n such that $n\eta^n = \mathcal{O}(1)$ and $(1-\eta)^{3-2\varepsilon}\log n \to 0$ as $n \to \infty$. Then, for sufficiently large n, the mapping $\alpha \mapsto \rho_n(\alpha)$ has a unique fixed point $\hat{\alpha}_n$ in $\mathscr{A}_n$ almost surely, such that $\rho_n(\hat{\alpha}_n) = \hat{\alpha}_n$. Furthermore, for any $\hat{\alpha}_n^{(0)} \in \mathscr{A}_n$, the sequence $\{\hat{\alpha}_n^{(m)}\}$ defined by (2.5) converges to $\hat{\alpha}_n$ almost surely as $m \to \infty$, provided n is sufficiently large.*

It is easy to see that $n\eta^n = n(1-\delta)^n = \mathcal{O}(1)$ implies $\delta n \to \infty$ because $(1-\delta)^n = (1-\delta n/n)^n \approx \mathrm{e}^{-\delta n}$. Therefore, to satisfy $n\eta^n = \mathcal{O}(1)$, it is required that the rate at which δ tends to zero be lower than that of n^{-1}. On the other hand, the condition that $\delta^{3-2\varepsilon}\log n \to 0$ is equivalent to the requirement that the convergence rate of δ be greater than that of $(\log n)^{-1/(3-2\varepsilon)}$, which automatically rules out the case of constant η. For any $v \in (0,1)$, the choice of $\delta = \mathcal{O}(n^{-v})$ satisfies all the conditions in Theorem 1.

Depending on how quickly δ tends to zero, different rates of weak convergence to normality can be established for the CM frequency estimator $\hat{\omega}_n$. Two useful cases are considered in the following.

Theorem 2. *Assume that the conditions in Theorem* 1 *are satisfied so that the CM algorithm defined by* (2.5) *converges to the fixed point* $\hat{\alpha}_n \in \mathscr{A}_n$. *Let* $\hat{\omega}_n$ *be the resulting frequency estimator obtained from* (2.6). *If* η *is chosen such that* $(1-\eta)^2 n \to \infty$ *and* $(1-\eta)^5 n \to 0$ *as* $n \to \infty$, *then* $(1-\eta)^{-3/2} n^{1/2}(\hat{\omega}_n - \omega_0) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \gamma^{-2})$ *as* $n \to \infty$.

The requirements in Theorem 2 can be satisfied by the choice of $1-\eta = \mathcal{O}(n^{-v})$ for any $v \in (\frac{1}{5}, \frac{1}{2})$. The required initial precision for the CM algorithm to converge can then be expressed as $\mathcal{O}(n^{-\varepsilon v})$ which can be made arbitrarily close to $\mathcal{O}(n^{-1/5})$ by choosing v close to $\frac{1}{5}$ and ε close to unity. Therefore any crude estimator with an initial guess $\hat{\alpha}_n^{(0)}$ of accuracy less than $\mathcal{O}(n^{-1/5})$ is accurate enough to initiate the CM algorithm and the algorithm is guaranteed to converge to the desired estimator with the asymptotic error rate (i.e., asymptotic standard error of $\hat{\omega}_n$ which takes the form $\mathcal{O}_{\mathrm{P}}(n^{-(1+3v)/2})$) at most $\mathcal{O}_{\mathrm{P}}(n^{-4/5})$.

Theorem 3. *Under the conditions in Theorem* 1, *if* η *is chosen such that* $(1-\eta)^2 n \to 0$ *as* $n \to \infty$, *then* $(1-\eta)^{-1/2} n(\hat{\omega}_n - \omega_0) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \gamma^{-1})$ *as* $n \to \infty$.

The conditions in Theorem 3 can be satisfied by the choice of $1-\eta = \mathcal{O}(n^{-v})$ for any $v \in (\frac{1}{2}, 1)$. With this choice, the resulting asymptotic standard error of $\hat{\omega}_n$ can be expressed as $\mathcal{O}_{\mathrm{P}}(n^{-1-v/2})$ and the required initial precision can be expressed as $\mathcal{O}(n^{-\varepsilon v})$. Clearly, as v takes values near unity, the error rate can be made arbitrarily close to $\mathcal{O}_{\mathrm{P}}(n^{-3/2})$ – the optimal error rate achieved by the PM and NLS methods.

The above theorems suggest a practical and computationally efficient procedure: start the CM algorithm with a constant η, use the resulting estimate to initiate the CM algorithm of Theorem 2, and the results, in turn, are sufficiently accurate to initiate the CM algorithm of Theorem 3. Thus, we have just outlined a single algorithm consisting of initial, intermediate, and final steps. In this way, our CM-based iterative procedure overcomes the computational difficulties associated with PM and NLS methods and is guaranteed to converge by the theoretical analysis summarized before. In terms of computational complexity, our procedure is also efficient in that it requires $m = \mathcal{O}(\log n)$ iterations in the initial step to improve the frequency estimation from an initial accuracy of $\mathcal{O}(1)$ to an accuracy good enough to serve as initializer for the intermediate step and it takes $\mathcal{O}(1)$ iterations in the intermediate and final steps. Since each iteration involves $\mathcal{O}(n)$ multiplications to compute the estimator, the total computational complexity equals $\mathcal{O}(n \log n)$, which is what FFT requires. But unlike the FFT method, the CM estimator has a nearly optimal accuracy of $\mathcal{O}(n^{-3/2})$ instead of a suboptimal accuracy of $\mathcal{O}(n^{-1})$ produced by the FFT method. In addition to the computational efficiency, in view of its recursive filtering scheme, this procedure can be implemented on line and easily updated when new observations come in unlike block estimation procedures such as Fourier-transform-based approaches.

Truong-Van (1990) and Quinn and Fernandes (1991) discussed some iterative procedures of frequency estimation which attain the optimal error rate of $\mathcal{O}(n^{-3/2})$. These algorithms, similarly to the PM and NLS methods, require initial values of precision $o(n^{-1})$. The algorithm discussed by Quinn and Fernandes (1991) can be viewed as a limiting case of the CM method with $\eta = 1$ [i.e., AR(2) filter without the regularization

Table 1
Simulated and Theoretical Standard Errors ($\phi = 0,\ \gamma = 1$)

η	$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = 700$	$n = 900$
η_1	1.698E-02	1.009E-02	5.701E-03	4.448E-03	3.864E-03	3.136E-03
LKY	1.387E-02	9.807E-03	5.662E-03	4.386E-03	3.707E-03	3.269E-03
η_2	6.100E-03	2.550E-03	8.063E-04	4.975E-04	3.598E-04	2.667E-04
Thm2	3.871E-03	2.004E-03	7.056E-04	4.343E-04	3.155E-04	2.485E-04
η_3	2.184E-03	7.007E-04	1.331E-04	6.632E-05	4.243E-05	2.899E-05
Thm3	9.844E-04	3.998E-04	9.584E-05	4.934E-05	3.186E-05	2.298E-05
η_4	1.815E-03	5.647E-04	1.112E-04	5.040E-05	2.965E-05	1.939E-05
Thm3	5.474E-04	2.004E-04	4.074E-05	1.942E-05	1.192E-05	8.283E-06
NLS	1.559E-03	5.513E-04	1.061E-04	4.931E-05	2.977E-05	2.041E-05

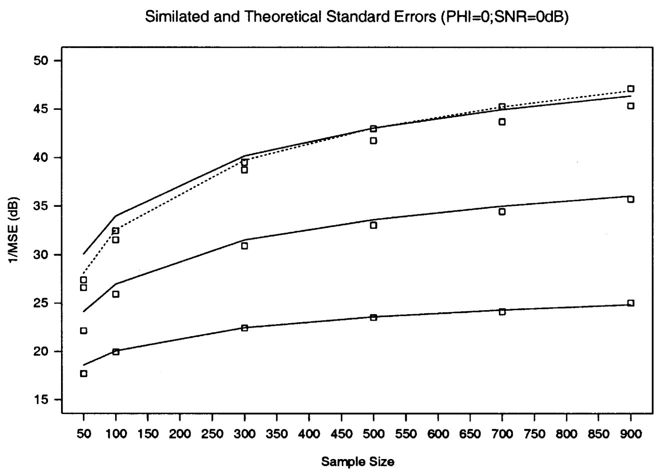

Fig. 4. Plot of 1/MSE (in decibels) as a function of η and n. Solid lines stand for theoretical results obtained from (bottom up) LKY ($\eta = \eta_1$), Thm2 ($\eta = \eta_2$), and Thm3 ($\eta = \eta_3$). Dashed line is the asymptotic Cramér–Rao lower bound (PM/NLS). Squares are simulation results based on 200 independent trials.

parameter]. As we have seen before, one of primary benefits of the CM algorithm lies in the flexible choice of η.

4. Numerical experiments

To confirm the theoretical findings in the previous section, we conducted some simulation experiments. The simulation results are presented in Table 1 and Fig. 4.

In these experiments, the noise $\{\varepsilon_t\}$ is white Gaussian, the amplitude β of the sinusoid equals unity, the phase ϕ equals zero, and the variance σ_ε^2 of the noise is chosen for each realization to attain the desired signal-to-noise ratio of $\gamma = 1$. The true frequency is $f_0 := \omega_0/(2\pi) = 0.21$. Presented in Table 1 are the root mean-squared errors (RMSEs) of the CM frequency estimator $\hat{f}_n := \hat{\omega}_n/(2\pi)$ based on 200 independent realizations of $\{y_t\}$ in which the phase is held fixed. For each sample size n, four values of η are used, namely $\eta_1 = 0.4$, $\eta_2 = 1 - n^{-0.3}$, $\eta_3 = 1 - n^{-0.6}$, and $\eta_4 = 1 - n^{-0.9}$.

For each value of η and each sample size n, the theoretical (asymptotic) standard error is also given in Table 1 immediately below the corresponding RMSE from the simulation. Note that the first value of η is a constant and away from unity. Therefore, the theoretical standard errors (labeled LKY) are obtained from the results of Li et al. (1994). The second and third values of η depend on n with the choice of v satisfying the conditions of Theorems 2 and 3, respectively, so that the theoretical standard errors can be obtained from these results. Similar remarks apply to the fourth value of η. In the last row of Table 1 are given the theoretical standard errors of the PM and NLS (or maximum likelihood) estimators whose asymptotic distribution coincides with the asymptotic distribution of the CM estimator with $\eta = 1$, namely $n^{3/2}(\hat{\omega}_n - \omega_0) \xrightarrow{\mathrm{D}} \mathcal{N}(0, 12\gamma^{-1})$ (Quinn and Fernandes, 1991).

Note that the same initial value $\hat{\omega}_n^{(0)} = 0.05 \times 2\pi$ is used in the simulation for all sample sizes, hence the initial precision is $\mathcal{O}(1)$. Starting with this initial value and with $\eta = \eta_1$, the CM algorithm employs the increasing values of η sequentially. As the value of η increases, the previous frequency estimates are used to initiate the CM algorithm with the next value of η. The convergence of this procedure is ensured theoretically by Theorem 1–3 and the results of Li and Kedem (1993a) and Li et al. (1994), and is also confirmed by the simulation. Furthermore, the RMSE of the final frequency estimates with $\eta = \eta_4$ closely approximates the theoretical standard error of the PM and NLS estimators, namely $12\gamma^{-1}n^{-3/2} = \mathcal{O}(n^{-3/2})$, which is also the asymptotic Cramér–Rao lower bound (CRLB) for frequency estimation.

Although the analytical analysis is carried out for single sinusoid, the CM algorithm can be applied to multiple sinusoids in colored noise under suitable conditions (Li and Kedem, 1994). Fig. 5 shows a simulated example to demonstrate this application. In this example, the time series contains two superimposed sinusoids whose frequencies are well separated with respect to the frequency resolution n^{-1} where $n = 128$. The additive noise $\{\varepsilon_t\}$ is the superposition of two independent random processes: a white noise process and an AR(2) process of the form $\xi_t + a_1\xi_{t-1} + a_2\xi_{t-2} = \zeta_t$, where $\zeta_t \sim \mathrm{WN}(0, \sigma_\zeta^2)$, $a_1 = -0.5$, $a_2 = 0.6$. The noise variances are adjusted for the given realization of $\{\varepsilon_t\}$ so that the SNR is equal to -8.94 dB for the sinusoid with frequency $f_1 = 19.5n^{-1}$ and -7.00 dB for the sinusoid with frequency $f_2 = 42.5n^{-1}$. The periodogram in Fig. 5(a) shows that the first frequency cannot be distinguished from the spurious peaks of the colored noise. It also shows that the peak locations of the periodogram do not necessarily provide accurate frequency estimates. The CM algorithm, as shown in Fig. 5(b), successfully converges to one of the signal frequencies, depending on the initial value. The final frequency estimates are as accurate as in the case of single sinusoid.

5. Proof of the theorems

According to (2.2) and (2.3), one can write

$$y_t(\alpha) + \eta^2 y_{t-2}(\alpha) = y_t + (1 + \eta^2)\alpha y_{t-1}(\alpha).$$

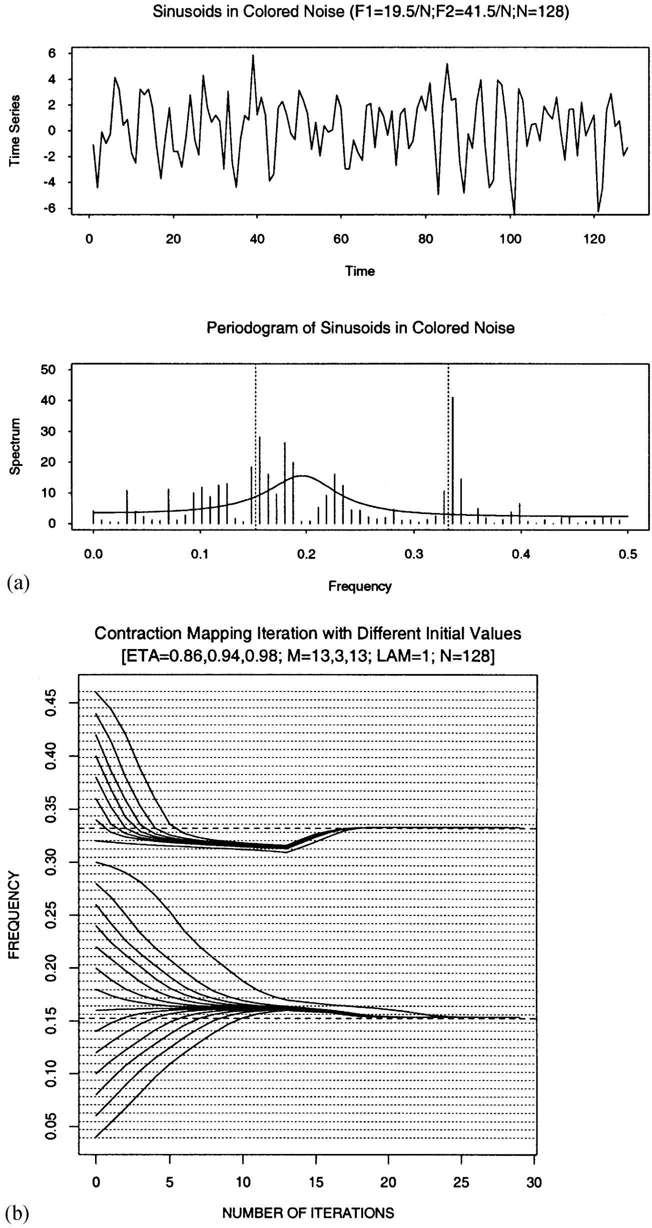

Fig. 5. (a) A time series and its periodogram which consists of two sinusoids in colored noise. Noise spectrum is superimposed on the periodgram, with vertical dotted lines indicating the signal frequencies. (b) Frequency estimates obtained from the CM algorithm with different initial values.

Therefore, $\rho_n(\alpha)$ in (2.4) can be expressed as

$$\rho_n(\alpha) = \alpha + (1+\eta^2)^{-1}\sin\lambda\,\frac{U_n(\lambda)}{V_n(\lambda)}, \tag{5.1}$$

where

$$V_n(\lambda) := \sin^2\lambda \sum_{t=1}^{n} y_{t-1}^2(\alpha), \tag{5.2}$$

$$U_n(\lambda) := \sin\lambda \sum_{t=1}^{n} y_t y_{t-1}(\alpha). \tag{5.3}$$

In these expressions λ is determined by α according to (2.3). In order to investigate the properties of the random mapping $\alpha \mapsto \rho_n(\alpha)$ in the neighborhood of α_0, it is necessary to understand the behavior of $V_n(\lambda)$ and $U_n(\lambda)$ in a neighborhood of λ_0, where $\lambda_0 \in (0,\pi)$ is determined by

$$\cos\lambda_0 = \frac{1+\eta^2}{2\eta}\alpha_0. \tag{5.4}$$

This is the main focus of the following section.

5.1. Preliminary results

Let Λ_n be the set of $\lambda \in (0,\pi)$ determined by (2.3) with $\alpha \in \mathscr{A}_n \subset \mathscr{A}$. Since $\eta \to 1$ and hence $\mathscr{A} \to (-1,1)$ as $n \to \infty$, it follows that α_0 becomes an interior point of $\mathscr{A}$ for large n. Further, since the size of $\mathscr{A}_n$ decreases with the increase of n, the interval $\mathscr{A}_n$, being a shrinking neighborhood of α_0, can be contained, for large n, in a closed subinterval of $\mathscr{A}$ which is independent of n. As a result, the interval Λ_n, with sufficiently large n, can be contained by a closed subinterval of $(0,\pi)$, say $\Lambda \subset (0,\pi)$, which are independent of n. This implies that any $\lambda \in \Lambda_n$ can be uniformly bounded away from 0 and π for large n.

The following propositions and corollaries describe some asymptotic characteristics of $V_n(\lambda)$ and $U_n(\lambda)$ as functions of $\lambda \in \Lambda_n$. The proofs of all these results are given in Song and Li (1997).

Proposition 1. *Let $V_n(\lambda)$ be defined by (5.2) with $\lambda \in \Lambda_n$ and $\alpha \in \mathscr{A}_n$. Assume that $\delta = 1-\eta \to 0$ as $n \to \infty$, but $n\eta^n = n(1-\delta)^n = \mathcal{O}(1)$. Then for $\varepsilon \in (1, \frac{3}{2})$, one obtains*

$$\begin{aligned} V_n(\lambda) = {} & \tfrac{1}{8}\beta^2\eta^{-2}\delta^{-2}n + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-3/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-1}n\log n) \\ & + (\lambda-\lambda_0)\{\mathcal{O}(\delta^{\varepsilon-4}n) + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-5/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-2}n\log n)\}, \end{aligned} \tag{5.5}$$

almost surely and uniformly in $\lambda \in \Lambda_n$ for sufficiently large n, and

$$\begin{aligned} & V_n(\lambda) - V_n(\lambda') \\ & \quad = (\lambda-\lambda')\{\mathcal{O}(\delta^{\varepsilon-4}n) + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-5/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-2}n\log n)\} \end{aligned} \tag{5.6}$$

almost surely and uniformly in $\lambda, \lambda' \in \Lambda_n$.

Proposition 2. *Let $U_n(\lambda)$ be defined by* (5.3). *If the conditions in Proposition* 1 *are satisfied, then one obtains*

$$\begin{aligned}U_n(\lambda) &= \mathcal{O}(\delta^{-1/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-1}\sqrt{n\log n}) + \mathcal{O}(\delta^{-3/2}\sqrt{\log n})\\ &\quad + (\lambda-\lambda_0)\{\tfrac{1}{4}\beta^2\delta^{-2}n + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-3/2}n\sqrt{\log n})\\ &\quad + \mathcal{O}(\delta^{\varepsilon-3}n) + \mathcal{O}(\delta^{-2}\sqrt{n\log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n})\},\end{aligned} \tag{5.7}$$

almost surely and uniformly in $\lambda\in\Lambda_n$ for sufficiently large n, and

$$\begin{aligned}U_n(\lambda) - U_n(\lambda') &= (\lambda-\lambda')\{\tfrac{1}{4}\beta^2\delta^{-2}n + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-3/2}n\sqrt{\log n})\\ &\quad + \mathcal{O}(\delta^{\varepsilon-3}n) + \mathcal{O}(\delta^{-2}\sqrt{n\log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n})\},\end{aligned} \tag{5.8}$$

almost surely and uniformly in λ, $\lambda'\in\Lambda_n$.

As a corollary to Proposition 1, we have the following results that are instrumental to the proof of Theorem 1.

Corollary 1. *Let the conditions in Proposition* 1 *be satisfied. Assume further that* $\delta^{3-2\varepsilon}\log n\to 0$. *Then, it follows that*

$$V_n(\lambda) = \tfrac{1}{8}\beta^2\eta^{-2}\delta^{-2}n\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{2\varepsilon-2})\},$$

$$V_n(\lambda_0) = \tfrac{1}{8}\beta^2\eta^{-2}\delta^{-2}n\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n})\},$$

$$V_n(\lambda)V_n(\lambda') = \tfrac{1}{64}\beta^4\eta^{-4}\delta^{-4}n^2\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{2\varepsilon-2})\},$$

$$V_n(\lambda) - V_n(\lambda') = (\lambda-\lambda')\mathcal{O}(\delta^{\varepsilon-4}n),$$

almost surely and uniformly in λ, $\lambda'\in\Lambda_n$ for sufficiently large n.

For the proof of Theorem 1, we also need the following corollary concerning $U_n(\lambda)$ that directly results from Proposition 2.

Corollary 2. *Assume that the conditions in Corollary* 1 *are satisfied. Then, it follows that*

$$U_n(\lambda) = \mathcal{O}(\delta^{\varepsilon-2}n),$$

$$U_n(\lambda_0) = \mathcal{O}(\delta^{-1/2}n\sqrt{\log n}),$$

$$U_n(\lambda) - U_n(\lambda') = (\lambda-\lambda')\tfrac{1}{4}\beta^2\delta^{-2}n\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\},$$

almost surely and uniformly in λ, $\lambda'\in\Lambda_n$ for large n.

We also need the following results to establish the asymptotic distribution of the CM frequency estimator $\hat{\omega}_n$.

Proposition 3. *Under the conditions of Proposition* 1, *it follows that*

$$U_n(\lambda_0) = W_{n,1} + W_{n,2} + \mathcal{O}_{\mathrm{P}}(\delta^2 n) + \mathcal{O}_{\mathrm{P}}(n^{1/2}) + \mathcal{O}_{\mathrm{P}}(\delta^{-1}),$$

where $W_{n,1}$ and $W_{n,2}$ are defined by

$$W_{n,1} := \frac{1}{2}\beta g(\zeta_0)\delta \sum_{t=1}^{n} \varepsilon_t(\eta^{n-t} - \eta^{t-1}) \sin(t\omega_0 + \phi), \tag{5.9}$$

$$W_{n,2} := \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda_0)\varepsilon_t\varepsilon_{t-j} \tag{5.10}$$

with $g(\lambda) := (1 - 2\eta\cos\lambda + \eta^2)^{-1}$ and $\zeta_0 := \lambda_0 - \omega_0$.

Proposition 4. *Let $W_{n,1}$ be defined by* (5.9). *If the conditions of Proposition* 3 *are satisfied, then $\delta^{3/2} W_{n,1} \xrightarrow{\mathrm{D}} \mathcal{N}(0, \frac{1}{8}\beta^2\sigma_\varepsilon^2)$ as $n \to \infty$.*

Proposition 5. *Let $W_{n,2}$ be defined by* (5.10). *If the conditions of Proposition* 3 *are satisfied, then $\delta^{1/2} n^{-1/2} W_{n,2} \xrightarrow{\mathrm{D}} \mathcal{N}(0, \frac{1}{4}\sigma_\varepsilon^4)$ as $n \to \infty$.*

5.2. Proof of Theorem 1

Equipped with Corollaries 1 and 2, we now turn to the proof of Theorem 1. The objective is to show that $\alpha \mapsto \rho_n(\alpha)$ is a *contractive mapping* in $\mathscr{A}_n$. This requires two inequalities [e.g., Stoer and Bulirsch, 1980, p. 251, Theorem 5.2.3]:

(a) There exists a constant $c \in (0, 1)$ such that

$$|\rho_n(\alpha) - \rho_n(\alpha')| \leqslant c\,|\alpha - \alpha'| \tag{5.11}$$

uniformly for all α, $\alpha' \in \mathscr{A}_n$.

(b) With the same constant c, it holds that

$$|\rho_n(\alpha_0) - \alpha_0| \leqslant (1 - c)\,\kappa\,(1 - \eta)^{\varepsilon}. \tag{5.12}$$

Of course, these inequalities need to hold almost surely for sufficiently large n.

5.2.1. Proof of Inequality (5.11)

From (5.1) we obtain

$$\rho_n(\alpha) - \rho_n(\alpha') = \alpha - \alpha' + R_n, \tag{5.13}$$

where

$$R_n := (1 + \eta^2)^{-1} \frac{J_1 + J_2 + J_3}{V_n(\lambda)V_n(\lambda')}$$

with

$$J_1 := (V_n(\lambda') - V_n(\lambda))U_n(\lambda)\sin\lambda',$$

$$J_2 := (U_n(\lambda) - U_n(\lambda'))V_n(\lambda)\sin\lambda',$$

$$J_3 := (\sin\lambda - \sin\lambda')U_n(\lambda)V_n(\lambda').$$

According to Corollaries 1 and 2, we can write

$$J_1 = (\lambda - \lambda')\mathcal{O}(\delta^{2\varepsilon-6}n^2),$$

$$J_2 = (\lambda - \lambda')\sin\lambda' \tfrac{1}{32}\beta^4\eta^{-2}\delta^{-4}n^2\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\},$$

$$J_3 = (\lambda - \lambda')\mathcal{O}(\delta^{\varepsilon-4}n^2)\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{2\varepsilon-2})\}.$$

To obtain the third expression, we also use the fact that $\sin\lambda - \sin\lambda' = (\lambda - \lambda')\cos\lambda''$, where λ'' lies between λ and λ'. Because Λ_n is contained in a closed subinterval of $(0, \pi)$, it follows that $\sin\lambda'$ can be bounded away from zero uniformly for all $\lambda' \in \Lambda_n$. Therefore, we can write

$$\begin{aligned} J_1 + J_2 + J_3 = {} & (\lambda - \lambda')\sin\lambda' \tfrac{1}{32}\beta^4\eta^{-2}\delta^{-4}n^2\{1 + \mathcal{O}(\delta^{-1}n^{-1}) \\ & + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\}. \end{aligned}$$

This result, combined with the expression for $V_n(\lambda)V_n(\lambda')$ in Corollary 1, leads to

$$R_n = (\lambda - \lambda')\sin\lambda' 2\eta^2(1+\eta^2)^{-1}\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\}.$$

Note that Lemma 9(b) of Song and Li (1997) provides an expression for $\lambda - \lambda'$ in terms of $\alpha - \alpha'$. Using this expression with the fact that $\alpha - \alpha' = \mathcal{O}(\delta^{\varepsilon})$, we can write

$$R_n = -\eta(\alpha - \alpha')\{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\}.$$

Substituting this result in (5.13) yields

$$\rho_n(\alpha) - \rho_n(\alpha') = C_n(\alpha - \alpha'), \tag{5.14}$$

where

$$C_n := \delta + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1}). \tag{5.15}$$

The proof is complete upon noting that $C_n \to 0$ almost surely and uniformly in α, $\alpha' \in \mathscr{A}_n$. □

5.2.2. Proof of Inequality (5.12)

Using (5.1) we can write

$$\rho_n(\alpha_0) - \alpha_0 = (1+\eta^2)^{-1}\sin\lambda_0 \frac{U_n(\lambda_0)}{V_n(\lambda_0)},$$

where λ_0 is defined by (5.4). According to Corollaries 1 and 2, we obtain

$$\frac{U_n(\lambda_0)}{V_n(\lambda_0)} = \frac{\mathcal{O}(\delta^{3/2}\sqrt{\log n})}{1 + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n})} = \mathcal{O}(\delta^{3/2}\sqrt{\log n}).$$

Because $\delta^{3/2-\varepsilon}\sqrt{\log n} \to 0$, it follows that $\mathcal{O}(\delta^{3/2}\sqrt{\log n}) = o(\delta^{\varepsilon})$. This implies that

$$\rho_n(\alpha_0) - \alpha_0 = o(\delta^{\varepsilon})$$

almost surely for large n. The proof is thus complete. □

5.3. Proof of Theorem 2

Let $\hat{\alpha}_n$ be the fixed point of $\rho_n(\alpha)$ in $\mathscr{A}_n$. Then, according to (5.14), we can write

$$\begin{aligned}\hat{\alpha}_n - \alpha_0 &= \rho_n(\hat{\alpha}_n) - \rho_n(\alpha_0) + \rho_n(\alpha_0) - \alpha_0 \\ &= C_n(\hat{\alpha}_n - \alpha_0) + \rho_n(\alpha_0) - \alpha_0.\end{aligned}$$

This implies that

$$\hat{\alpha}_n - \alpha_0 = (1 - C_n)^{-1}\{\rho_n(\alpha_0) - \alpha_0\}.$$

Using (5.1), we obtain

$$\hat{\alpha}_n - \alpha_0 = (1 + \eta^2)^{-1}(1 - C_n)^{-1}\sin\lambda_0 \frac{U_n(\lambda_0)}{V_n(\lambda_0)}.$$

Note that Corollary 1 ensures that

$$\delta^2 n^{-1} V_n(\lambda_0) \to \tfrac{1}{8}\beta^2$$

almost surely as $n \to \infty$. Note also that $C_n \to 0$ almost surely and $\sin\lambda_0 \to \sin\omega_0$. Therefore, by Slutsky's theorem,

$$\delta^{-3/2} n^{1/2}(\hat{\alpha}_n - \alpha_0) \sim 4\beta^{-2}\sin\omega_0 \delta^{1/2} n^{-1/2} U_n(\lambda_0) \tag{5.16}$$

asymptotically, so that it suffices to consider the distribution of $\delta^{1/2} n^{-1/2} U_n(\lambda_0)$.

Note that $W_{n,1} = \mathcal{O}_{\mathrm{p}}(\delta^{-3/2})$ by Proposition 4. Therefore, under the assumption that $\delta^2 n \to \infty$ and $\delta^5 n \to 0$, one can show from Proposition 3 that $\delta^{1/2} n^{-1/2} U_n(\lambda_0)$ has the same asymptotic distribution as $\delta^{1/2} n^{-1/2} W_{n,2}$. This result, combined with Proposition 5, leads to

$$\delta^{-3/2} n^{1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathrm{D}} \mathscr{N}(0, \gamma^{-2}\sin^2\omega_0).$$

The proof is complete upon noting that, by the delta method, $\delta^{-3/2} n^{1/2}(\hat{\omega}_n - \omega_0)$ has the same asymptotic distribution as $\delta^{-3/2} n^{1/2}(\hat{\alpha}_n - \alpha_0)(\sin\omega_0)^{-1}$. □

5.4. Proof of Theorem 3

Using a similar argument as in the proof of Theorem 2, we can show that

$$\delta^{-1/2} n(\hat{\alpha}_n - \alpha_0) \sim 4\beta^{-2}\sin\omega_0\, \delta^{3/2} U_n(\lambda_0) \tag{5.17}$$

asymptotically, so that it suffices to consider the distribution of $\delta^{3/2} U_n(\lambda_0)$.

To that end, we use the expression of $U_n(\lambda_0)$ given by Proposition 3 together with the asymptotic normality of $W_{n,1}$ ensured by Proposition 4. According to these results and the assumption that $\delta^2 n \to 0$, it is easy to see that $\delta^{3/2} U_n(\lambda_0)$ has the same asymptotic distribution as $\delta^{3/2} W_{n,1}$ so that $\delta^{3/2} U_n(\lambda_0) \xrightarrow{\mathrm{D}} \mathscr{N}(0, \frac{1}{8}\beta^2\sigma_\varepsilon^2)$. Combining this result with (5.17) yields

$$\delta^{-1/2} n(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathrm{D}} \mathscr{N}(0, \gamma^{-1}\sin^2\omega_0).$$

Again, an application of the delta method finishes the proof. □

6. Concluding remarks

In this paper we have considered the contraction-mapping (CM) method with variable regularization parameter for efficient frequency estimation, and provided a simple algorithm to overcome difficulties associated with PM and NLS methods. The CM algorithm is based on an iterative filtering idea in which the estimated first-order autocorrelation of the filtered process contracts to a fixed point that defines the frequency estimator. We have established, for the first time, the crucial connection between the accuracy of the initial guess required for convergence in the fixed point iteration and the precision of the CM estimator as the fixed point of the iteration; we have quantified the asymptotic relationship between the initial guess and the final CM estimator, together with the limiting distributions and almost sure convergence of the fixed point; and we have constructed a single unified algorithm adaptable to possibly poor initial values without requiring separate procedures such as FFT to provide initial values. It is shown that the CM algorithm, endowed with an adaptive regularization parameter, can accommodate possibly poor initial values of precision $\mathcal{O}(1)$ and converge to a final estimator whose precision is arbitrarily close to the optimal $\mathcal{O}(n^{-3/2})$. Although the presentation is focused on the single-frequency case, the CM method can be generalized to the multiple frequency case in two different ways: (a) If the frequencies are well separated, the single-frequency CM algorithm can be applied directly. Depending on the initial value, the CM algorithm will converge to one of the frequencies. (b) If the frequencies are closely spaced, it is more appropriate to employ the multivariate extension of the CM algorithm discussed in Li and Kedem (1993b, 1994). Analytical investigation of the initial and final accuracies of the multivariate CM algorithm is an ongoing research.

Except for Theorem 2 and Proposition 5, all the results in this paper hold under the assumption that

$$\varepsilon_t = \sum_{j=0}^{\infty} \psi_j z_{t-j}, \tag{6.18}$$

where $\{\psi_j\}$ is absolutely summable with $\psi_0=1$ and $\{z_t\}$ is a stationary and ergodic martingale difference sequence with respect to $\mathscr{F}_t=\sigma(\{\varepsilon_s\colon s\leqslant t\})$ such that $E\{z_t^2|\mathscr{F}_{t-1}\}=\sigma^2$ almost surely and $E\{z_t^4\}<\infty$. For discussion on these conditions, see An et al. (1983) and Quinn and Fernandes (1991). Under these more general assumptions, to obtain limiting distribution results similar to Theorem 2 requires us to look at partial sum process of the form $\sum_{t=1}^{n}\varepsilon_t\sum_{j=1}^{t}w_j(n,\eta)\,\varepsilon_{t-j}$ where $w_j(n,\eta)$ are possibly random weights. To the best of the authors' knowledge, the existing martingale theory cannot be applied directly to such processes to obtain a limiting distribution (Chan and Wei, 1988). This is an interesting open problem for future research.

In the special case of white Gaussian noise, it deserves to discuss an interesting class of so-called signal/noise subspace methods of frequency estimation such as MUSIC, ESPRIT and MODE (e.g., Stoica and Nehorai, 1989; Stoica and Söderström, 1991; Roy and Kailath, 1989; Stoica and Sharman, 1990). These promising frequency estimators are based on the eigendecomposition methods of the sample covariance matrix and they all involve a design parameter m which is directly related to the dimension of the sample covariance matrix. This parameter m, if allowed to vary with the sample size n, can be viewed as playing a similar role as our variable regularization parameter η. For fixed m, the best convergence rate is known to be $n^{-1/2}$, which implies that the estimates are statistically inefficient as compared to the optimal rate $n^{-3/2}$. Many empirical studies and theoretical analyses suggest that the accuracy of these methods can be improved considerably by a suitable choice of m for the given sample size (e.g., Stoica and Nehorai, 1989). This raises the interesting question as to whether it would be possible for these methods to achieve the optimal rate $n^{-3/2}$ by varying m. No theoretical results have been reported that answer this question. In addition, many other important issues remain to be investigated in order to compare these methods with the CM procedure on an equal footing. These issues include, for example, (a) how to handle the computational complexity associated with finding the eigenvalues/vectors of the sample covariance matrix of huge dimension (the computational cost is proportional to m^3)? (b) how to quantify the relationship between the accuracy of the initial guesses and the precision of the final frequency estimates when iterations are involved?

Finally as we mentioned in the introduction, there are many other related methods in the engineering literature that are based on the idea of iterative filtering. An advantage of iterative filtering, as compared with eigen-decomposition based methods such as MUSIC, ESPRIT and MODE, is that it can be easily implemented to track time-varying frequencies. The adaptive notch filtering (ANF) algorithms considered by Nehorai (1985) and Tichavský and Händel (1995) are such examples. As pointed out by Li and Kedem (1994, 1998), the ARMA parametrization in these ANF algorithms produces inconsistent frequency estimates when the bandwidth parameter is fixed. The CM method discussed in this paper overcomes the inconsistency problem with an appropriately parametrized filter. More general discussion on this issue can be found in Li and Kedem (1993a, 1994, 1998). As far as analytical theory is concerned, it is unknown whether or not the ANF algorithms would have similar convergence and efficiency properties as the CM algorithm if the ANF bandwidth parameter is allowed to vary with the sample size, although it is so suggested by simulation results. A conjecture is that the ANF algorithms, when implemented without the forgetting mechanism, be asymptotically equivalent to the CM algorithm for estimating constant frequencies if the ANF bandwidth parameter tends to unity sufficiently rapidly as the sample size grows.

Acknowledgements

We wish to thank the Editor, an Associate Editor and two referees for their helpful comments on a previous version of this paper. The research of the first author was supported in part by FSU First-Year Assistant Professor Award.

References

An, H., Chen, Z., Hannan, E.J., 1983. The maximum of the periodogram. J. Multivariate Anal. 13, 383–400.
Chan, N.H., Wei, C.Z., 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. Ann. Statist. 16, 367–401.

Dragoševiċ, M.V., Stankoviċ, S.S., 1989. A generalized least squares method for frequency estimation. IEEE Trans. Acoust. Speech Signal Process. 37, 805–819.

Hannan, E.J., 1971. Nonlinear time series regression. J. Appl. Probab. 8, 767–780.

Kay, S.M., 1984. Accurate frequency estimation at low signal-to-noise ratio. IEEE Trans. Acoust. Speech Signal Process. 32, 540–547.

Kay, S.M., 1988. Modern Spectral Estimation: Theory and Application. Prentice-Hall, Englewood Cliffs, NJ.

Kedem, B., 1994. Time Series Analysis by Higher-Order Crossings. IEEE Press, Piscataway, NJ.

Li, T.H., 1996. Discrimination of time series by parametric filtering. J. Amer. Statist. Assoc. 91, 284–293.

Li, T.H., 1997. Time series characterization, Poisson integral and robust divergence measures. Technometrics 39, 357–371.

Li, T.H., Kedem, B., 1993a. Strong consistency of the contraction mapping method for frequency estimation. IEEE Trans. Inform. Theory 39, 989–998.

Li, T.H., Kedem, B., 1993b. Asymptotic analysis of a multiple frequency estimation method. J. Multivariate Anal. 46, 214–236.

Li, T.H., Kedem, B., 1994. Iterative filtering for multiple frequency estimation. IEEE Trans. Signal Process. 42, 1120–1132.

Li, T.H., Kedem, B., 1998. Tracking abrupt frequency changes. J. Time Ser. Anal. 19, 69–82.

Li, T.H., Kedem, B., Yakowitz, S., 1994. Asymptotic normality of the sample autocovariances with an application in frequency estimation. Stochastic Process. Appl. 52, 329–349.

Nehorai, A., 1985. A minimum parameter adaptive notch filter with constrained poles and zeros. IEEE Trans. Acoust. Speech Signal Process. 33, 983–996.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press, San Diego, CA.

Quinn, B.G., Fernandes, J.M., 1991. A fast efficient technique for the estimation of frequency. Biometrika 78, 489–497.

Rao, C.R., Zhao, L.C., 1993. Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals. IEEE Trans. Signal Process. 41, 1461–1464.

Rice, J.A., Rosenblatt, M., 1988. On frequency estimation. Biometrika 75, 477–484.

Roy, R., Kailath, T., 1989. ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. 37, 984–995.

Song, K.S., Li, T.H., 1997. Asymptotic theory of a frequency estimator. Technical Report, M917, Florida State University. Also available at `http://pstat.ucsb.edu/~thl/`.

Stoer, J., Bulirsch, R., 1980. Introduction to Numerical Analysis. Springer, New York.

Stoica, P., Moses, R.L., Friedlander, B., Söderström, T., 1989. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. IEEE Trans. Acoust. Speech Signal Process. 37, 378–392.

Stoica, P., Nehorai, A., 1988. Performance analysis of an adaptive notch filter with constrained poles and zeros. IEEE Trans. Acoust. Speech Signal Process. 36, 911–919.

Stoica, P., Nehorai, A., 1989. MUSIC, maximum likelihood, and Cramer-Rao bound. IEEE Trans. Acoust. Speech Signal Process. 37, 720–741.

Stoica, P., Sharman, K.C., 1990. Maximum likelihood methods for direction-of-arrival estimation. IEEE Trans. Acoust. Speech Signal Process. 38, 1132–1143.

Stoica, P., Söderström, T., 1991. Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies. IEEE Trans. Signal Process. 39, 1836–1847.

Tichavský, P., Händel, P., 1995. Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. IEEE Trans. Signal Process. 43, 1116–1127.

Truong-Van, B., 1990. A new approach to frequency analysis with amplified harmonics. J. Roy. Statist. Soc. Ser. B 52, 347–366.

Walker, A.M., 1971. On the estimation of a harmonic component in a time series with stationary independent residuals. Biometrika 58, 21–36.